

# Eedi コンペ<sup>o</sup>参戦記

kaerururu

from

terekaerumasahmet

# 目次

- ✓ 自己紹介
- ✓ コンペ参戦記
- ✓ まとめ

# 自己紹介



The image shows a Kaggle profile page for a user named 'kaerururu'. The profile includes a main profile picture of a black frog, a smaller profile picture of a blue frog with a crown, and a bio that identifies the user as an ML Engineer at a Japanese startup company in Yokohama, Japan. The user has been on Kaggle for 6 years and is a 'Competitions Grandmaster' with a rank of 283 out of 205,309. Below the bio, there are navigation tabs for 'About', 'Competitions (61)', 'Datasets (224)', 'Code (60)', 'Discussion (142)', 'Followers (200)', and 'Following (41)'. The 'Kaggle Achievements' section displays four achievement cards: 'Competitions Grandmaster' (Rank 283), 'Datasets Contributor' (Rank 173), 'Notebooks Master' (Rank 1,358), and 'Discussions Expert' (Rank 276).

kaerururu

ML Engineer at Japanese startup company  
Yokohama, Kanagawa, Japan  
Joined 6 years ago · last seen in the past day

Competitions Grandmaster  
283 of 205,309

About Competitions (61) Datasets (224) Code (60) Discussion (142) Followers (200) Following (41)

### Kaggle Achievements

Achievement	Rank	Highest Ever
Competitions Grandmaster	283	173
Datasets Contributor	1,358	117
Notebooks Master	276	103

名前: かえるるる

仕事: MLE

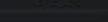
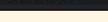
趣味: 🎧, ☕, kaggle

kaggle: Competitions Grandmaster

NLP、画像系のコンペが好き

Discussion メダルちまちま集めるのにハマってます

# 参加コンペ (抜粋)

	<b>PetFinder.my Adoption Prediction</b> How cute is that doggy in the shelter? Featured · Code Competition · 2023 Teams · 6 years ago	1/2023	
	<b>PetFinder.my - Pawpularity Contest</b> Predict the popularity of shelter pet photos Research · Code Competition · 3537 Teams · 3 years ago	5/3537	
	<b>Eedi - Mining Misconceptions in Mathematics</b> Predict affinity between misconceptions and incorrect answers (distractors) in multiple-choice questi... Featured · Code Competition · 1448 Teams · 25 days ago	7/1448	
	<b>Freesound Audio Tagging 2019</b> Automatically recognize sounds and apply tags of varying natures Research · Code Competition · 880 Teams · 6 years ago	7/880	
	<b>Feedback Prize - English Language Learning</b> Evaluating language knowledge of ELL students from grades 8-12 Featured · Code Competition · 2854 Teams · 2 years ago	13/2854	
	<b>Child Mind Institute - Detect Sleep States</b> Detect sleep onset and wake from wrist-worn accelerometer data Featured · Code Competition · 1877 Teams · a year ago	15/1877	
	<b>Rainforest Connection Species Audio Detection</b> Automate the detection of bird and frog species in a tropical soundscape Research · 7143 Teams · 4 years ago	16/7143	
	<b>LEAP - Atmospheric Physics using AI (ClimSim)</b> Simulate higher resolution atmospheric processes within E3SM-MMF, a climate model supported by 1... Research · 693 Teams · 6 months ago	19/693	
	<b>HuBMAP - Hacking the Human Vasculature</b> Segment instances of microvascular structures from healthy human kidney tissue slides. Research · Code Competition · 1021 Teams · a year ago	33/1021	
	<b>SIIM-ISIC Melanoma Classification</b> Identify melanoma in lesion images Featured · 3308 Teams · 4 years ago	37/3308	

- ・ 6年半で 61個のコンペ参加
  - ・ 22枚のメダル
- ・ NLP コンペが比較的得意
  - ・ w2v 時代 - deberta 時代 - LLM

# 計算環境

Eedi コンペ参加時の計算環境

- ・ 自宅マシン
  - ・ TitanRTX x2 48GB (写真→)
- ・ クラウド(自腹)
  - ・ Lambda Cloud H100 インスタンス (80GB)



かっこいい!

# コンペ概要

# タスク概要

- ✓ 数学の多肢選択問題が与えられる
- ✓ 不正解の背景にある誤解がどのようなものであるかを予測
- ✓ 誤解ラベルの候補 (2,587件) が用意されていて、この中から選択
- ✓ [kuto-san のまとめ記事](#) がわかりやすい

White Rose Maths

$5 \times 4 + 6 \div 2 =$

A	B	C	D
23	13	25	35

© White Rose 2017

「四則演算の順番に従わず、左から右に計算する」という誤解が隠れている

# データ概要

{train/test}.csv

- ✓ QuestionId に対して、複数の Answer と MisconceptionId が紐づいていた
- ✓ 横持ちのデータを縦持ちに変換して利用するのが主流だった (train\_long)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	QuestionId	ConstructId	ConstructName	SubjectId	SubjectName	Correct Answer	QuestionText	AnswerAText	AnswerBText	AnswerCText	AnswerDText	MisconceptionAId	MisconceptionBId	MisconceptionCId	MisconceptionDId
2	0	856	Use the order of operations to carry out calculations involving powers	33	BIDMAS	A	$\sqrt[3]{3 \times 2 + 4 - 5}$ Where do the brackets need to go to make the answer equal $\sqrt[3]{13}$ ?	$\sqrt[3]{3 \times (2+4) - 5}$	$\sqrt[3]{3 \times 2 + (4-5)}$	$\sqrt[3]{3 \times (2+4-5)}$	Does not need brackets				1672
3	1	1612	Simplify an algebraic fraction by factorising the numerator	1077	Simplifying Algebraic Fractions	D	Simplify the following, if possible: $\sqrt{\frac{m^2+2m-3}{m-3}}$	$\sqrt{m+1}$	$\sqrt{m+2}$	$\sqrt{m-1}$	Does not simplify	2142	143	2142	

# データ概要

misconception\_mapping.csv

✓ 2,587件の Misconception

✓ しかし、38% は欠損 (983件)

csv はスプシで EDA してます

	A	B	C	D	F	G
1	MisconceptionId	MisconceptionName				
2		0 Does not know that angles in a triangle sum to 180 degrees				
3		1 Uses dividing fractions method for multiplying fractions				
4		2 Believes there are 100 degrees in a full turn				
5		3 Thinks a quadratic without a non variable term, can not be factorised				
6		4 Believes addition of terms and powers of terms are equivalent e.g. $a + c = a^c$				
7		5 When measuring a reflex angle, gives the acute or obtuse angle that sums to 360 instead				
8		6 Can identify the multiplier used to form an equivalent fraction but does not apply to the numerator				
9		7 Believes gradient = change in y				
10		8 Student thinks that any two angles along a straight line are equal				
11		9 Thinks there are 180 degrees in a full turn				
12		10 Believes duration can be read from a timetable, rather than calculating from start and end times				

# 評価指標と提出方法

- ✓ Mean Average Precision @25
- ✓ 各質問-誤答のペアにつき、誤解の候補を25個提出できる
- ✓ 予測の順位が大事

```
QuestionId_Answer, MisconceptionId
```

```
1869_A,1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

```
1869_B,1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

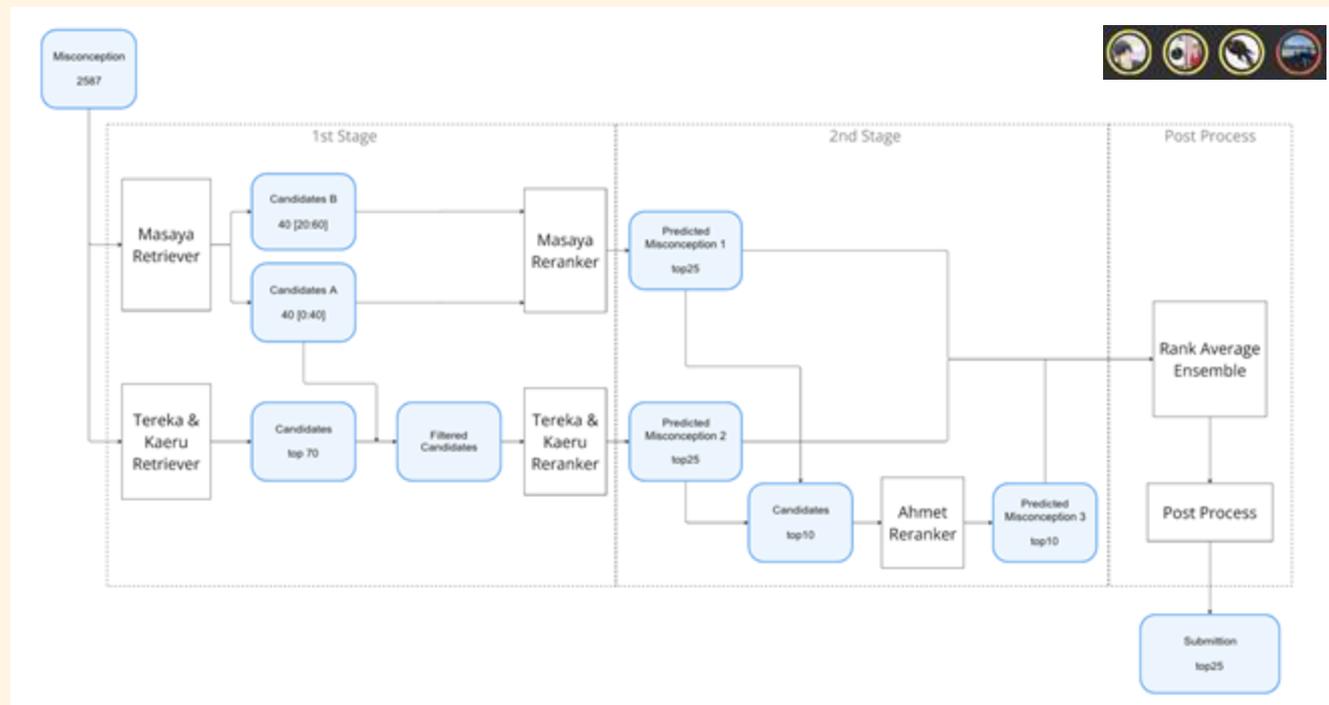
```
1869_C,1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

```
1870_B,1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

```
...
```

# 7th place solution

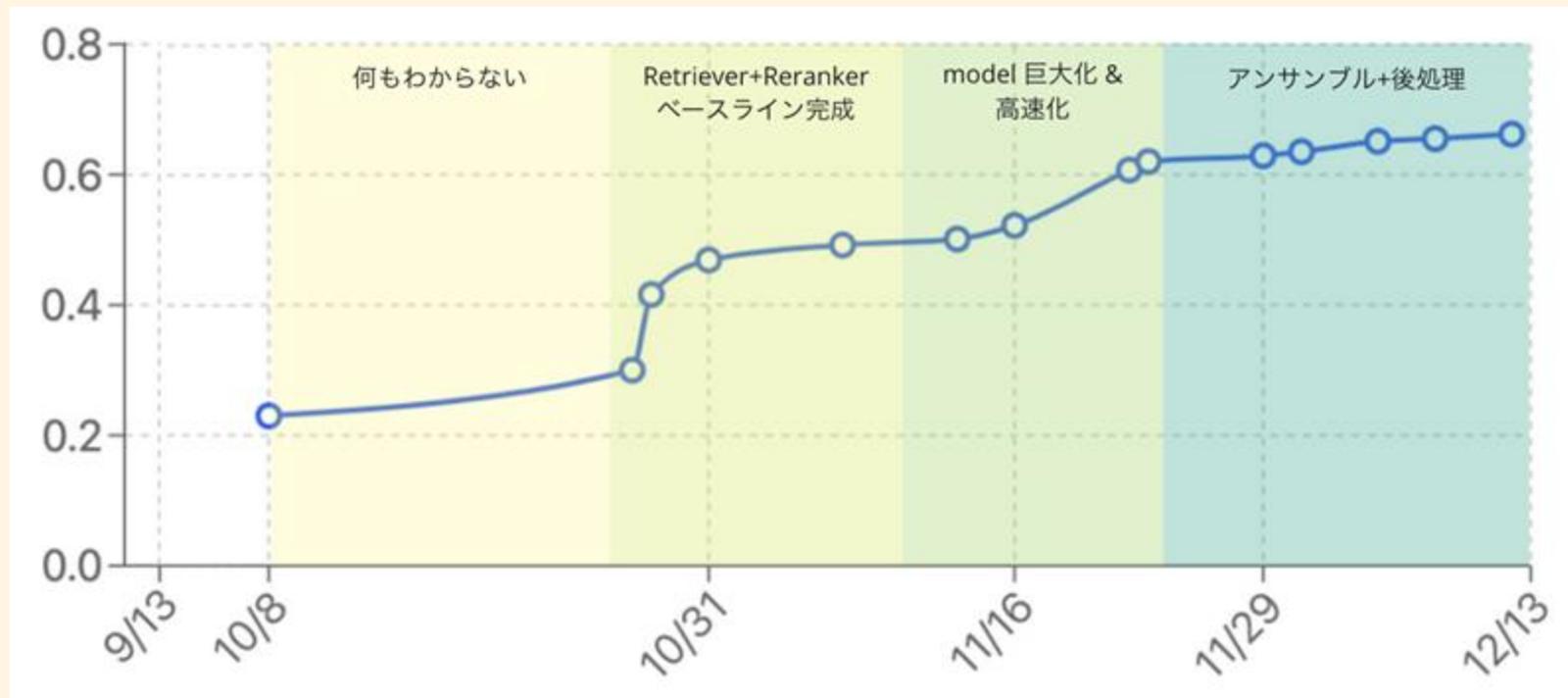
チームメンバーそれぞれのパイプラインの Rank Average + 後処理



# コンペ参戦記

# はじめに

期間ごとに、考えていたことや取り組んでいたことについてお話しします！



# 何もわからない期

## やったこと

- sinchir0-san の公開 notebook で始める
  - [tfidf retriever + debertav3 reranker](#) と [BGE retriever](#)
- sinchir0-san の BGE retriever が base の [takaito-san のスコア](#) が超えられない
  - → 何もわからなくなる
- そもそも transformers の Trainer の挙動がわからずキャッチアップ
- EDA で misconception の欠損に気づく



# 何もわからない期

## 考えていたこと

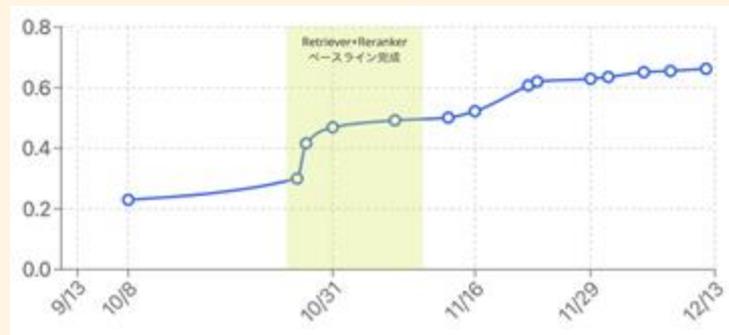
- Zero-shot Qwen2.5 32B 複数回実行で 0.36 超える公開 notebook 登場
  - → もしや LLM の finetune 必要なのでは...?
- 過去問で素振りしよう
  - かねてから LLM finetune をどこかでやっておきたかった
  - 評判のよかった [atmaCup17 の 1st place solution](#) にしよう
- misconception の欠損対策は後でやろう



# Retriever + Reranker に気づく

## 考えていたこと ①

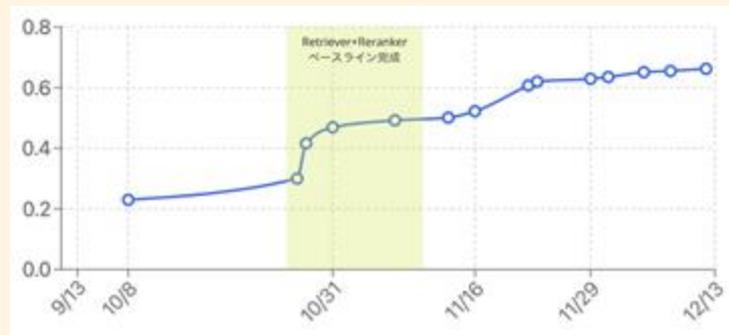
- baseline notebook に retriever + reranker があるのに、スコア高めの公開 notebook は retriever only model ばかりだった
- retriever は公開モデルで固定して reranker を gemma2 を試そう



# Retriever + Reranker に気づく

## やったこと ①

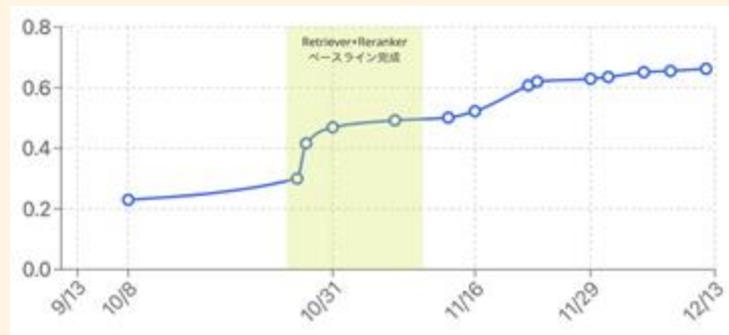
- retriever は公開モデルで固定して reranker として gemma2 の finetune
  - retriever の候補を負例、正解ラベルを正例として「どちらが誤解として適切か? yes, no で答えよ」の NSP で学習
  - 2B と 9B を試して、9B が当たる (0.30 → 0.416 🏆)
- 公開 notebook でハイスコア retriever が更新される度に、差し替え
- unsloth, sfairXC など色々なモデルを試す
- 0.492 (491min) まで上がる



# Retriever + Reranker に気づく

## 考えていたこと ②

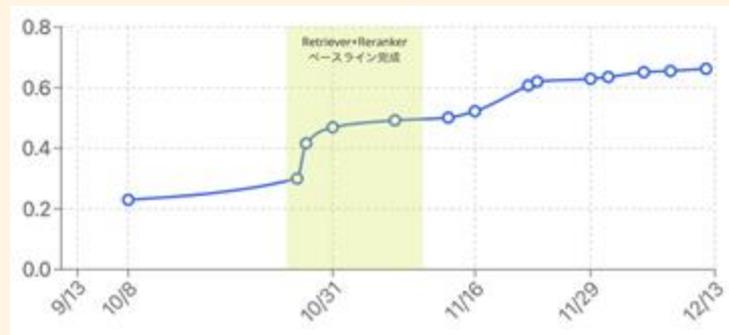
- ・ reranker の qlora tuning は時間も GPU も必要
- ・ 一人では厳しそう、チームマージしたい
- ・ チームマージに備えて、自分の推論パイプラインの高速化をしておかないと
  - ・ 過去コンペサーベイしよう
  - ・ 関西 kaggler 会参加者の X の眩きから vllm で推論すると早くなるらしい



# Retriever + Reranker に気づく

## やったこと ②

- lmsys など他の LLM 系のコンペの解法をサーベイ
- retriever を cuda:0 と cuda:1 とで並行実行するように修正
- 現在の推論コードを vllm で書き換える & 精度をある程度 keep する検証
- チームマージ打診
  - 睡眠コンペ#1 で一緒に戦った tereka-san に声をかけ、快諾してもらう



# model の巨大化、高速化

## やったこと ①

- ・ チームマージ
  - ・ 欠損への対応は tereka-san がやってくれた
- ・ tereka-san が reranker で qwen2.5 32B の実験を開始、少量データでの学習で kaeru の gemma2 9B の当時ベストスコアを超える (0.501 → 0.522)
- ・ kaeru はひたすら推論コードの blas up
  - ・ vllm による高速化と共通部分の流用



# model の巨大化、高速化

## チームで話していたこと ①

- ・ 当時の金圏が 0.6 超え
- ・ 一部データで 0.522 なら、全データなら期待できるのではないか



# model の巨大化、高速化

## やったこと ②

- qwen2.5 32B + 全データで学習で金圏へ (0.522 → 0.607)
- enable\_prefix\_cache パラメータによる高速化
- rerank 候補増やして (0.607 (260min) → 0.620 (401min))
- 更なるチームマージ



# model の巨大化、高速化

## チームで話していたこと ②

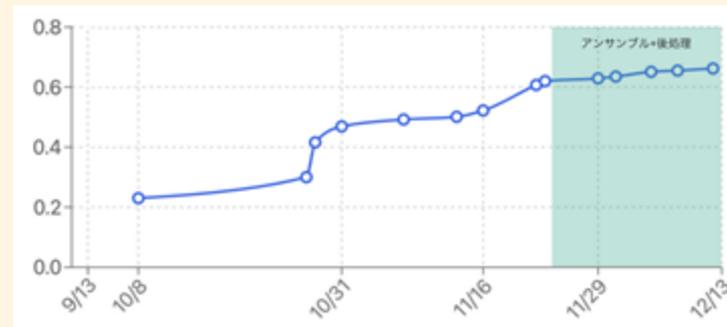
- ・ qwen2.5 72B も一部 topN件のみならいけるかも
  - ・ → private はよかったが public 伸びず



# Ensemble

## 考えていたこと

- enable\_prefix\_cache により推論時間大幅 save
- チームマージにより 3パターンの金圏解法
  - terekaeru pipeline
  - masaya-san pipeline
  - ahmet-san pipeline
- どうやってアンサンブルする？



# Ensemble

## やったこと

- Rank average によるアンサンブル (0.625 → 0.635)
  - 予測値 (test\_preds) の代わりに順位を使用
- 閾値で filtering
  - pipelineA の preds が低い QId\_Answer を pipelineB の推論対象から除く
  - 精度を keep しながら時間を save
- 浮いた時間で全員分の pipeline を 9h に詰め込めた！

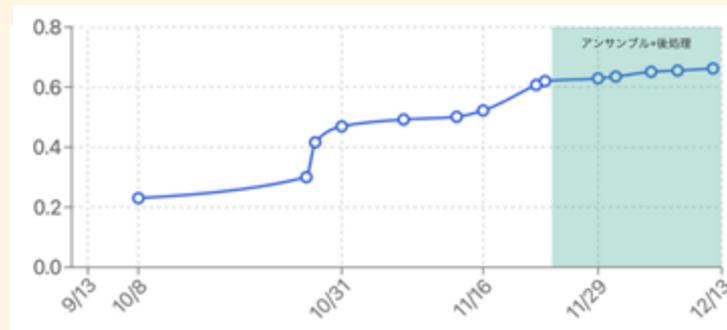
QuestionId_Answer	MisconceptionId	rank
str	str	f64
"1869_B"	"2306"	1.0
"1869_B"	"1345"	0.5
"1869_B"	"2518"	0.333333
"1869_B"	"1005"	0.25
"1869_B"	"1507"	0.2
...	...	...
"1871_D"	"1098"	0.047619
"1871_D"	"1521"	0.045455
"1871_D"	"1177"	0.043478
"1871_D"	"340"	0.041667
"1871_D"	"2151"	0.04

QuestionId_Answer	MisconceptionId	rank
str	str	f64
"1869_B"	"1345"	1.0
"1869_B"	"1005"	0.5
"1869_B"	"2488"	0.333333
"1869_B"	"1507"	0.25
"1869_B"	"706"	0.2
...	...	...
"1871_D"	"758"	0.047619
"1871_D"	"2211"	0.045455
"1871_D"	"2243"	0.043478
"1871_D"	"2562"	0.041667
"1871_D"	"188"	0.04



# 後処理 by Ahmet-san

- ・ 「すべての train に含まれていない誤解 (未知の誤解) は、少なくとも1つのテストセット行と一致する」という仮定のもと、未知の誤解の予測値を定数倍 (0.655 → 0.662)
  - ・ 3rd place だった人はこれ一本で最終日2日前くらいに public 1st に
- ・ private test には未知の誤解が答えの問題が多かったらしい
- ・ 意図的に未知データを作っていた host としては与えていない未知データに対する汎化性能が欲しい (それはそう)



まとめ

# まとめ

- ・ 9h に可能な限り推論を詰め込む
- ・ LB 上位の人が何をしているか想像して、戦略を立てる
- ・ 時間をかける

# 9h に可能な限り推論を詰め込む

まだまだアンサンブルは強い

- ・ LLM コンペは自由度が高く多様な解法が生まれやすい
- ・ kaggle 推論環境のリソース制約がきつい (16GB×2... え?)

まずは計測が基本

- ・ kaggle ではサブの実行時間が表示されない
- ・ カレーさんの sub 時間計測スクリプトを kaggle notebook で実行
  - ・ <https://zenn.dev/currypurin/scraps/47d5f84a0ca89d>
  - ・ サブが成功/失敗するまでバックグラウンドで実行してくれる

# 9h に可能な限り推論を詰め込む

## 様々な高速化テクニック

- ・ vllm の利用
- ・ vllm の `enable_prefix_cache` パラメータ
- ・ データを半分ずつにして `cuda:0` と `cuda:1` で並行実行

## 複数ステージ推論の場合、後段の候補を絞る

- ・ 全データ推論で `timeout` になる場合でも、TopN件だけを別のパイプラインで推論して並び替える

# LB 上位が何をしているか想像して、戦略を立てる

## 戦略を立てる

- ・ 欠損対策, Retriever, Reranker, 後処理どこが差分になっている?

## Public LB 1st になった時に考えていたこと

- ・ 32B モデルのアンサンブルで public 1st になった
- ・ 72B モデルを学習している人はいなさそう
  - ・ → ここまでやり切ったら差をつけられるかも!

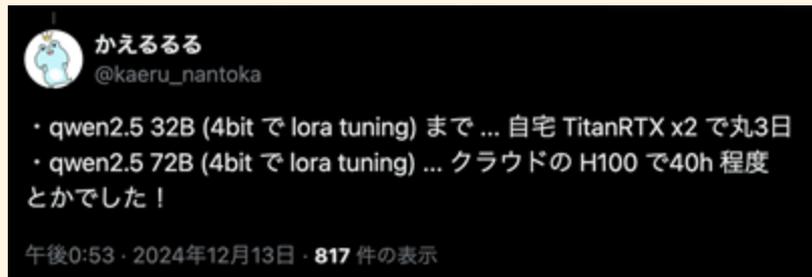
# 時間をかける

LLM 系のコンペはそもそも時間がかかる

- ・ 写真は pair-wise の Reranker の学習にかかった時間
- ・ 1 week チャレンジなどは超上級者向け

時間をかける ≠ ずっと kaggle に張り付く

- ・ とはいえ kaggle 漬け生活とはいかないことも
- ・ タスクランナーを活用し、複数の実験を一度に回す
  - ・ sh で 5fold 実行、定量評価、重みの kaggle に push までを自動で
  - ・ 外出の予定などがあっても実験が動いていると安心！



ご清聴ありがとうございました！